The Fourier Integral—A Basic Introduction

E. A. GUILLEMIN†

TO THE ENGINEER or applied scientist, the use of the Fourier integral, like any other analysis technique, is a means to an end. That is to say, it provides a method of analysis which is effective in obtaining solutions to certain kinds of physical problems. Strictly in this sense, the end result alone justifies the means, and it is wholly unnecessary that the method or any interpretive reasoning pertinent to it be subject to separate justification on any grounds other than a purely mathematical inquiry into its soundness. It is not necessary, for example, that the interpretive reasoning bear a direct relation to the nature of the physical problem and thus be required to make good physical sense. If we replace a periodic function representing some physical quantity by a sum of sinusoids because this expedient places in evidence a simple way of solving the particular problem in hand, it is improper to question the separate physical existence of each sinusoidal component. From an analysis standpoint this question is wholly irrelevant. If an *imagined* separate physical existence of each sinusoid supports one's ability to acquire a facile use of the Fourier method, then such quasi-physical interpretation is justified; but in such situations one should never lose sight of the fact that this practice is equivalent to playing a game of "make-believe," since any other attitude may seriously interfere with the process of understanding this important and useful mathematical tool.

Having thus defined what our attitude toward the Fourier method should be from an analysis point of view, we are free to indulge in whatever interpretive games of "make-believe" we may find expedient. Thus when the resultant interference pattern of a sum of sinusoids having appropriate amplitudes, frequencies, and time phases, is used to replace a given physical excitation or response function, we may find it useful to attribute a physical reality to each sinusoidal component, even though it is only the resultant which can truthfully be so interpreted. Through thus attributing to our mathematical mechanism a greater degree of physical reality than it deserves, we are in many situations in a position to use physical and intuitive reasoning in conjunction with the formal mathematical rules of procedure to achieve a more meaningful and certain understanding of the validity of a contemplated operation. In other words, such interpretations often permit the exercise of common sense to guide us in a subtle situation that might, in terms of purely mathematical procedures alone, be difficult to resolve. We shall reemphasize the importance of this attitude in the final paragraph of this introduction.

Visualization of an interference pattern created by the Fourier representation of a time function is, in general, simpler in the case of periodic time functions, for which this representation takes the form of a series rather than that of an integral; and so it is customary to lead gradually to the integral representation by starting with a discussion of periodic functions. Actually it turns out that one arrives in this way at a representation of nonperiodic functions (defined over an infinite interval) immediately, without making the usual transition from the series to an integral. In fact, the Fourier series representation yields directly the tool for handling functions that are sampled at regular intervals, whether they be periodic or not; and a rather lucid way of making a transition to the Fourier integral is to allow the sampling interval to become smaller and smaller until the pertinent function is defined in a continuous manner.

In order to introduce our topic from this point of view, it is expedient to regard the mathematical aspect of the Fourier series at the outset as amounting essentially to a change-of-variable process. Suppose we consider this process applied to the independent variable in the time function $f(t)$ so as to replace t by a new independent variable ω, and denote the function in terms of this new variable as $a(\omega)$. The Fourier series representation enables us to compute $f(t)$ from $a(\omega)$ or vice versa; and in this representation one of the independent variables is continuous, while the other assumes only discrete values. More specifically, if t is the continuous variable then $f(t)$ is periodic, while the sampled function $a(\omega)$ is not. The Fourier representation in this case may be written

$$f(t) = \frac{1}{\omega_0} \sum_{\omega=-\infty}^{\infty} a(\omega) e^{j\omega t} \Delta\omega \tag{1}$$

and

$$a(\omega) = \frac{1}{t_0} \int_{-t_0/2}^{t_0/2} f(t) e^{-j\omega t} dt, \tag{2}$$

in which the discrete variable ω is specifically defined by writing

$$\omega = k\omega_0, \quad \text{with} \quad k = 0, +1, +2, \cdots \tag{3}$$

and the sampling interval for the function $a(\omega)$ is denoted by

$$\Delta\omega = \omega_0, \tag{4}$$

while the fundamental period of the periodic function is

$$t_0 = 2\pi/\Delta\omega = 2\pi/\omega_0. \tag{5}$$

In the pair of relations (1) and (2) we now interchange the symbols ω and t, and correspondingly, ω_0 and t_0; then we replace ω by $-\omega$, write $b(\omega)$ for $f(-\omega)$ and $h(t)$ for $a(t)$, all of which is permitted because it amounts merely

† Massachusetts Institute of Technology, Cambridge 39, Mass.

to a trival change in notation.[1] This gives

$$b(\omega) = \frac{1}{t_0} \sum_{t=-\infty}^{\infty} h(t)e^{-i\omega t}\Delta t \tag{6}$$

and

$$h(t) = \frac{1}{\omega_0} \int_{-\omega_0/2}^{\omega_0/2} b(\omega)e^{j\omega t}d\omega \tag{7}$$

in which t is now a discrete variable defined by

$$t = kt_0, \qquad k = 0, \pm 1, \pm 2, \cdots ; \tag{8}$$

the sampling interval for the time function is

$$\Delta t = t_0, \tag{9}$$

and the fundamental period of the continuously defined function $b(\omega)$ is

$$\omega_0 = 2\pi/\Delta t = 2\pi/t_0. \tag{10}$$

The sampled time function $h(t)$ is arbitrary (except for the usual restrictions imposed by the Fourier theorem), while $b(\omega)$ is a periodic function.

In the more conventional presentation of (1) and (2), the function $a(\omega) = a(k\omega_0)$ is written a_k and is designated a *Fourier coefficient*. In the series (1), the quantities ω_0 and $\Delta\omega$ are cancelled, and we have

$$f(t) = \sum_{k=-\infty}^{\infty} a_k e^{jk\omega_0 t} \tag{11}$$

with

$$a_k = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} f(t)e^{-jk\omega_0 t}dt. \tag{12}$$

The equivalent representation in (1) and (2) stresses primarily two things: The functional character of the Fourier coefficient $a_k = a(k\omega_0) = a(\omega)$; and the duality existing between the pair of functions $f(t)$ and $a(\omega)$ with which this representation is concerned.

This duality manifests itself in the interchangeability of the roles which these two functions play and of their respective characteristics; namely, that one function is *continuously defined* while the other one is *sampled*. Although this duality has long been recognized, still when the idea of a sampled function was first introduced into the communications picture, its spectrum representation was in some instances carried out by unnecessarily round-about methods simply because the forms (11) and (12) in which the Fourier series representation is usually stated obscured the fact that the answer was obvious by inspection. The outward appearance or form of a mathematical relationship, although generally speaking of secondary importance, is sometimes of greater significance than one might expect.

This kind of an approach to the Fourier series representation has a further advantage in that it immediately suggests a wider field of usefulness for this mathematical tool than merely the handling of circuit analysis problems involving periodic excitation forces, although its introduction into the usual electrical engineering curriculum is appropriately done through the medium of this or some similar problem. Motivation for the use of Fourier methods in this connection is based primarily upon two things: (1) the fact that circuit analysis in the sinuosidal steady state is simply and compactly expressed in terms of the impedance concept, and (2) the fact that the Fourier theorem enables one to express arbitrary functions as sums of steady sinusoids. Putting these two thoughts together leads us to see that the simplicity and circumspection which the impedance concept seems initially to provide only for the consideration of steady sinusoidal excitation functions appears also to be available in essentially unaltered form for the consideration of any practical kind of excitation function, regardless of form or character. This thought is indeed a strong motivation, and leads naturally to the *time-domain frequency-domain* concept, in terms of which one regards the Fourier method not merely as a harmonic analysis procedure, but rather as a transformation method whereby time as an independent variable is replaced by frequency, and the differential equations expressing system equilibrium are replaced by equivalent algebraic equations.

Not only is the analysis side of a problem simplified through these interpretive means, but the measurement aspects likewise are made easier, since the spectrum function is more readily measurable than the time response. The knowledge that the two are uniquely related, and having the means for carrying out this relationship, add immeasurably to one's ability to cope with problems where excitation and response functions are apt to have rather arbitrary forms.

It is essential in this connection to point out that use of these concepts leads also to some rather subtle difficulties, and that the consequences of introducing the Fourier theorem into engineering analysis are not all to be found on the credit side of our ledger. These difficulties stem essentially from two circumstances, namely, that we cannot escape from becoming involved in some way with an *approximation problem*, and that this problem is aggravated by having two domains (time and frequency) to consider, since errors of approximation in these two domains need somehow to be effectively correlated if the potential value in the use of Fourier methods is to be fully realized.

Thus in the replacement of a periodic time function by a Fourier series, practical matters usually prohibit the consideration of an infinite number of terms, and

[1] Replacing ω by $-\omega$ is done in order to preserve for the forms (6) and (7) the conventional orientation of the complex frequency plane so that integration around the left half-plane singularities of the spectrum function corresponds to $t > 0$ for the time function. In carrying out this sign change in (2), the effect upon the limits of integration should not be overlooked. Regarding the question as to whether changes in notation of this sort are permissive, I usually find it necessary to remind my students that the choice of symbols for variables is anyone's privilege, and he can dispose of it according to his own pleasure. Thus he may replace ω or t by the picture of an inverted cow, if he wants to. Questions of mathematical justification are not even remotely involved here.

the question immediately arises as to the error that will be committed through replacing the infinite series by a partial sum. Questions of convergence, as well as those concerned with the approximation properties of the partial sums, need to be clarified in this connection. These matters become especially acute when Fourier methods are used, as they inescapably must be, in the problem of network synthesis for a prescribed transient response. Here we are usually faced with an approximation problem in both the time and frequency domains, and hence with the necessity of correlating their respective errors.

In these considerations certain inherent properties of the pair of relations (1) and (2) or of their equivalents (6) and (7) or (11) and (12) are important. Thus, whenever the partial sum of a Fourier series is used in connection with a practical problem, it is essential to be aware of the fact that the resulting approximation automatically has the character that renders the mean square error a minimum. That is to say, if we denote the partial sum of (11) by

$$s_n(t) = \sum_{-n}^{n} a_k e^{jk\omega_0 t}, \tag{13}$$

define the error function

$$\epsilon(t) = f(t) - s_n(t), \tag{14}$$

and ask for that set of coefficients a_k which will make the integral of ϵ^2 over a period $t_0 = 2\pi/\omega_0$ a minimum, then we arrive at formula (12). Hence the mean-square-error criterion is seen to be an *implicit* property of the Fourier series. Whenever we use Fourier means in an approximation procedure, we are automatically applying the mean-square-error criterion, whether we like it or not. If we do not like it, we must use some finite trigonometric polynomial other than the partial sum of a Fourier series.

A second important implicit property of the Fourier series is tied directly to the mean-square-error criterion. This is the Gibbs phenomenon which appears at points where $f(t)$ has discontinuities. It is demonstrated through using the Fourier series to express a function $f(t)$ in terms of itself. The analysis is a little easier (and just as appropriate) if we first make the transition from a series to an integral, because an integral between finite limits is more readily evaluated than the corresponding sum. Let us consider for this transition the pair of expressions (6) and (7), and let the sampling interval $\Delta t = t_0$ for the time function approach the differential value dt. In order to have the associated spectrum function remain finite it is necessary to define it as $g(\omega) = t_0 b(\omega)$. The sum (6) becomes an integral and we have

$$g(\omega) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt \tag{15}$$

and

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\omega) e^{j\omega t} d\omega \tag{16}$$

which are referred to as the *Fourier integrals*.

A study of the approximation properties of the partial sum of a Fourier series is equivalent in the integrals (15) and (16) to inquiring as to how well (16) with $g(\omega)$ substituted from (15) approximates $h(t)$ if the infinite limits on the integral (16) are replaced by finite limits. After interchanging the order of the two integrations, this process yields

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(\xi) \int_{-\omega_c}^{\omega_c} e^{j\omega(t-\xi)} d\omega d\xi \tag{17}$$

in which ω_c is the finite upper limit used for the evaluation of the integral (16). If we define the function

$$D(t - \xi) = \frac{1}{2\omega_c} \int_{-\omega_c}^{\omega_c} e^{j\omega(t-\xi)} d\omega = \frac{\sin \omega_c(t - \xi)}{\omega_c(t - \xi)} \tag{18}$$

then (17) may be written

$$h(t) = \frac{\omega_c}{\pi} \int_{-\infty}^{\infty} h(\xi) D(t - \xi) d\xi. \tag{19}$$

The function (18) is referred to as the *Dirichlet kernel*, and the integral (19) defines the so-called *convolution* of the function $h(t)$ with this kernel. This convolution integral has many interesting and useful properties which are helpful in its evaluation. First, it may be shown that the arguments ξ and $t-\xi$ can be interchanged, which is equivalent to stating that the convolution of two functions is strictly a mutual process. Second, we can differentiate one of the two functions and integrate the other without affecting the result, and this process can be repeated as often as we please. If one function is differentiated or integrated and the other is not, then we obtain the derivative or integral of $h(t)$, and we can regain $h(t)$ by subsequent integration or differentiation.

As an illustration of this flexible attitude, suppose we assume that $h(\xi)$ is a unit step function, and that our object is to study the approximation properties of the Fourier series or integral representation in the vicinity of the discontinuity. Since the derivative of h is a unit impulse, and since the convolution of any function with a unit impulse equals that same function, we see that $h(t)$ in (19) becomes the integral of the $\sin x/x$ function in (18). The net area enclosed by the function (18) is π/ω_c, and so it is clear that the function (19) ultimately changes by unity as it should, but this change proceeds in a gradual and ripply fashion with a 9 per cent overswing near the point of discontinuity in $h(t)$. Thus the familiar Gibbs phenomenon is obtained.

The slope of this resulting $h(t)$ function at its half-value point is ω_c/π times the value of $D(0)$ which is unity. Hence a reasonable definition of rise time is π/ω_c seconds which, then, is roughly the time required for the output of a lowpass filter with cutoff at ω_c to follow abrupt changes in its input (apart from the delay caused by the average slope of the associated phase function).

Although the Fourier theorem justifies the relationship between the pair of functions in (15) and (16) only so long as ω and t remain real variables, one may regard g as a function of $j\omega$ rather than of ω, and, writing p for $j\omega$, let p assume complex values. The extended function $g(p)$ evidently coincides with the Fourier transform $g(j\omega)$ on the j axis of the complex frequency plane, and hence, by the identity theorem pertinent to functions of a complex variable, $g(p)$ must be identical with the analytic continuation of $g(j\omega)$ into the complex plane. The resulting integrals are written

$$g(p) = \int_{-\infty}^{\infty} h(t)e^{-pt}dt \tag{20}$$

and

$$h(t) = \frac{1}{2\pi j}\int_{-\infty}^{\infty} g(p)e^{pt}dp. \tag{21}$$

Methods of complex integration may now be used for the evaluation of the inverse transformation (21).

Unlike the conventional Laplace integrals, (20) and (21) do not involve the tacit assumption that the time function is zero for $t<0$. They are applicable whether $h(t)$ is or is not zero for negative values of t, and hence include the usually defined Laplace transforms as a special case.

If in (15) and (16) we separate $h(t)$ into its even and odd parts by writing

$$h(t) = h_e(t) + h_0(t), \tag{22}$$

and if $g(\omega)$ is written in the explicit complex rectangular form

$$g(\omega) = g_1(\omega) + jg_2(\omega), \tag{23}$$

then it follows rather straightforwardly that

$$g_1(\omega) = \int_{-\infty}^{\infty} h_e(t)\cos\omega t dt, \quad g_2(\omega) = \int_{-\infty}^{\infty} h_0(t)\sin\omega t dt, \tag{24}$$

from which we note, incidentally, that the real and imaginary parts of $g(\omega)$ are respectively even and odd functions of ω. The substitution of (23) into (16) and the separation into two integrals yields the respective inverse transforms to those stated in (24), namely,

$$h_e(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_1(\omega)\cos\omega t d\omega,$$

$$h_0(t) = \frac{-1}{2\pi}\int_{-\infty}^{\infty} g_2(\omega)\sin\omega t d\omega. \tag{25}$$

Here, as in (24), the limits of integration may be extended only over positive values of the independent variable and the result multiplied by two, since the integrands are even functions.

If the time function, as with Laplace transforms, is tacitly assumed zero for $t<0$, then it is clear that h_e and h_0 cancel for $t<0$, and hence must be equal for $t>0$. Therefore such an $h(t)$ for $t>0$ is given by either twice h_e or twice h_0 for $t>0$, we have from (25)

$$h(t) = \frac{2}{\pi}\int_0^{\infty} g_1(\omega)\cos\omega t d\omega = \frac{-2}{\pi}\int_0^{\infty} g_2(\omega)\sin\omega t d\omega. \tag{26}$$

Since the impulse response of a passive network must be dead before the advent of the excitation (assumed to occur at $t=0$), we see that the real or imaginary part of thefresponse function $g(\omega)$ is sufficient to characterize this impulse response.

This result, which reveals the real part of a system function to be sufficient for the characterization of a network, is independently obtainable by showing that any rational function which is analytic in the right half of the complex frequency plane can uniquely be constructed from its real or imaginary part defined on the j axis. The uniqueness claim in this statement can be met if we include impulses in the real part corresponding to possible j axis poles of the rational function.[2]

The real-part sufficiency which is contained in these statements has far-reaching practical value for both analysis and synthesis applications, since it is a means for avoiding the complex character of network response functions and the many computational annoyances that derive therefrom. In a transfer function, for example, one is rid of having to consider whether a minimum-phase or nonminimum-phase function is involved. The real part of the system function and the time response are both real functions of real variables, and hence the time and frequency-domain characterizations are more homogeneous.[3]

In concluding this very condensed introductory survey of the essential features of the Fourier method, I would like to stress the importance of having a sound understanding of the physical behavior of the systems with which the engineer deals and of the broad conceptual implications of the various purely mathematical subtleties which are involved in some of the evaluation procedures. Theorems and limit processes involving epsilons and deltas are essential, of course, in order to round out the mathematical soundness; but the engineer who relies on these artificial crutches instead of on his common sense and intuition to guide him in the practical use of mathematics will be so slowed down by ponderous equipment and the ghastly sound of strange phrases that he will be unable to do any creative work or get any pleasure out of doing the things that the mathematical theorems claim are not permitted. Mathematical theorems are written like legal documents and are equally useless, in most practical situations. I much prefer a statement that is inaccurate or incomplete as to detail, but meaningful and easy to understand. I am willing that my common sense should guide me past the pitfalls. Anyone who uses the Fourier method and its several related branches of analysis with this attitude will find the experience both enjoyable and rewarding.

[2] "Computational Techniques which Simplify the Correlation between Steady-State and Transient Response of Filters and Other Networks," RLE-MIT Technical Report No. 268, Sept. 2, 1953, or *Proc. Natl. Elect. Conf.*, vol. 9, pp. 513–532; February, 1954.

[3] The mathematical basis for the exploitation of this attitude is given in the RLE-MIT Technical Report No. 233, June 4, 1952, "On Basic Existence Theorems; III Theoretical Considerations on Rational Fraction Expansions for Network Functions."